



## *Proyectos de bibliotecas digitales de matemáticas soportados por la Cellule MathDoc*

Thierry Bouche

Cellule MathDoc & institut Fourier,  
Université Joseph Fourier, Grenoble

II Jornada DOCUMAT  
Madrid, 26 de Noviembre de 2009

NUMDAM



cedram



- 1 La documentación matemática
- 2 La Cellule MathDoc y sus servicios
- 3 La biblioteca virtual de matemáticas (WDML)
- 4 El proyecto EuDML

# La documentación matemática

## Particularidades

- La documentación matemática queda pertinente a largo plazo (Euler 1999)
- Teoremas antiguos no están superados por resultados nuevos, sino que sirven de base para ellos (Richelot 2004)
- Es válida en su *totalidad*, creando una red de dependencias y referencias (Corona bug)
- Es útil para otras ciencias, de manera *asíncrona* (Weber crypto)

# *La documentación matemática*

## *Necesitas*

Entonces, la documentación matemática necesita

- Custodia con precaución: validación (científica (Corona bug)), indexación precisa y preservación (GDZ Spr. Zbl MR)
- Acceso a largo plazo (Galois 1828)

# *La documentación matemática*

## *La biblioteca matemática de referencia*

Es decir que necesitamos una (¡por lo menos!) biblioteca, idealmente

- completa
- al día
- bien organizada y manejada
- muy abierta
- fácil de usar para todos

**Papel** Vale? (red de bibliotecas, préstamo interbibliotecario, suministro de documentos, catálogos colectivos, bases de recensiones (MR/ZM). . .)

**Electrónico** Un sueño difícil de realizar (WDML)

⇒ **EuDML (2010-2012)** proyecto piloto (contenido digitalizado y nacido digital desde bibliotecas digitales «nacionales» y editoriales, búsqueda de documentos con fórmulas, accesibilidad mejorada)

# *La documentación matemática*

## *La biblioteca matemática de referencia*

Es decir que necesitamos una (¡por lo menos!) biblioteca, idealmente

- completa
- al día
- bien organizada y manejada
- muy abierta
- fácil de usar para todos

**Papel** Vale? (red de bibliotecas, préstamo interbibliotecario, suministro de documentos, catálogos colectivos, bases de recensiones (MR/ZM). . .)

**Electrónico** Un sueño difícil de realizar (WDML)

⇒ **EuDML (2010-2012)** proyecto piloto (contenido digitalizado y nacido digital desde bibliotecas digitales «nacionales» y editoriales, búsqueda de documentos con fórmulas, accesibilidad mejorada)

# *La documentación matemática*

## *La biblioteca matemática de referencia*

Es decir que necesitamos una (¡por lo menos!) biblioteca, idealmente

- completa
- al día
- bien organizada y manejada
- muy abierta
- fácil de usar para todos

**Papel** Vale? (red de bibliotecas, préstamo interbibliotecario, suministro de documentos, catálogos colectivos, bases de recensiones (MR/ZM). . .)

**Electrónico** Un sueño difícil de realizar (WDML)

⇒ **EuDML (2010-2012)** proyecto piloto (contenido digitalizado y nacido digital desde bibliotecas digitales «nacionales» y editoriales, búsqueda de documentos con fórmulas, accesibilidad mejorada)

# La documentación matemática

## Cifras

Se estiman tales cifras describiendo el tamaño del corpus escrito de matemáticas publicado según la tradición científica occidental desde Euclides:

- 3 millones de textos en < 100 millones de páginas
- 100,000 textos nuevos publicanse cada año
- < 20 % antes de 1900
- > 50 % después de 1950
- 80 % artículos de revistas, 10 % capítulos en libros colectivos, 10 % libros, monografías, tesis, etc.
- 600 revistas vivas dedicadas a matemáticas, 2000 revistas o periódicos *con* artículos matemáticos
- ¿15 millones de páginas ya digitalizadas? (Rehmann WDML)



# Cellule MathDoc

## Misiones

La *Cellule de coordination documentaire nationale pour les mathématiques* (MathDoc) es una pequeña «unidad mixta de servicios» CNRS-UJF en Grenoble desde 1995

### ● Misiones

- 1995 Coordinación nacional por la documentación matemática (junto con RNBM  $\simeq$  DOCUMAT francés)
  - Ayuda técnica y servicios electrónicos para bibliotecarios y matemáticos
  - Colaboración con el *Zentralblatt-MATH* (buscador/servidor)
- 2000 Digitalización de las principales revistas francesas
- 2005 Desarrollo de herramientas para la publicación electrónica de revistas académicas

# Cellule MathDoc

## Proyectos

Pequeño grupo con 3 informáticos, 1-2 documentalistas, 2 matemáticos

- **Proyectos principales**

**1996** CFPM: Catálogo colectivo de las bibliotecas de la red francesa RBNM (última actualización con ayuda de Silvia Molero de Tenerife en 2009)

**1997** EDBM: Web interfaz y buscador por la base de recensioni ZMATH

**2002** NUMDAM I: digitalización pilota de 5 periódicos

**2005** CEDRAM, v. 1: edición electrónica de 3 revistas

- LiNuM, mini-DML, Gallica-Math, NUMDAM II;

**2007-** CEDRAM, v. 2: 5 revistas, 4 seminarios, metadatos MathML/ $\text{\LaTeX}$ .

# NUMDAM

## Síntesis

«Programa de biblioteca digital para el archivo y la difusión de los fondos de revistas académicas de matemáticas»

**Archivo:** Metadatos y ficheros gráficos con gran precisión

**Digitalización:** Escaneo integral desde la primera página hasta la última de cada cuaderno (TIFF 600 dpi mono)

**Nacido digital** Adquisición del PDF de cada artículo proporcionado por su editorial.

**Difusión:** en [www.numdam.org](http://www.numdam.org)

**Acceso** Un fichero de varias páginas por cada artículo, accesible después de una *ventana móvil*

**Navegación** Consulta o busca libre de todos los metadatos  
Muchas enlaces facilitan la navegación

# NUMDAM

## Colecciones

### 2000-2004 I Fase

6 periódicos (4 importantes revistas de matemáticas puras, 1 serie de monografías brevas, 1 serie de actas de congreso)

**7 500 artículos, 210 000 p.**

### 2004-2007 II Fase

17 revistas (desde física matemática hasta estadística, y desde Nîmes en 1810 hasta Amsterdam, Pisa en XXI siglo. . .)

28 seminarios (de los cuales: Bourbaki [I.H.P.], Cartan [É.N.S.], Leray [Collège de France], . . .)

**+ 20 000 artículos, + 375 000 p.**

### 2008-2010 III Fase

Revistas mas antiguas, tesis 1918-1940, revistas de mat. aplicadas, estadísticas. . .

Acabamos de adquirir artículos nacido digital desde CEDRAM, Elsevier, Springer, EDP Sciences, próximamente project Euclid.

**+ 25 000 artículos, + 300 000 p.?**

**10 % de nuestro contenido nacido digital**

# NUMDAM

## *Enlaces de todos tipos*

- Cada artículo tiene enlaces hasta MR, ZM, y otras bases de datos pertinentes (SPS)
- Había una base de resúmenes del Séminaire de probabilités de Strasbourg desarrollada por MathDoc antes de la digitalización: hemos enlazado ella con NUMDAM (SPS)
- Cada artículo tiene su bibliografía si es formalizada al final de dicho artículo. Cada artículo citado tiene enlaces hasta MR, ZM si se ha encontrado ahí, y también enlaces hasta el texto integral si se ha encontrado en NUMDAM, o gracias a la mini-DML.

# NUMDAM

## Apuntes

- Cuando dos artículos se siguen en la misma página, «limpiamos» los ficheros de cada uno (BSMF, Gergonne, NAM)
- Proporcionamos índices impresos en los seminarios (GAU)
- ¡Aquella revista publica algunos artículos en trozos: los unimos de nuevo! (AFST)

# NUMDAM

## Novedades

- Tenemos muchas clases de enlaces específicos para algunas revistas: Errata <sup>(AIF)</sup> ; Artículo siguiente <sup>(Schwartz)</sup> ; Complementos <sup>(NAM)</sup> ; Problemas/soluciones en *Annales de Gergonne* y *Nouvelles annales de mathématiques* <sup>(NAM)</sup>
- Cuando un artículo es proporcionado por una editorial, pongamos un enlace DOI (o PURL) <sup>(CEDRAM, Elsevier, Springer, EDP Sciences)</sup>
- Tenemos muchas imágenes que no difundimos en nuestra web (paratexto como índices, páginas de títulos, láminas fuera de texto, etc.). En algunos casos, proponemos para descargar desde el índice electrónico
  - páginas preliminares y finales <sup>(BSMF)</sup>
  - el volumen completo en un fichero (con tales imágenes en gris o color si lo necesitan) <sup>(AMPA, NAM)</sup>

# CEDRAM

## Síntesis

*«Centro de difusión de revistas académicas de matemática»*

**Soporte:** Herramientas para mejorar la calidad de producción de revistas independientes de investigación en matemática

**Alojamiento:** Servidores dedicadas a la edición electrónica

**Archivo:** Traslado de los ficheros PDF/XML en NUMDAM para acceso a largo plazo, de los ficheros de producción en un archivo de preservación



# CEDRAM

## Colecciones

- *Annales de la faculté des sciences de Toulouse, mathématiques* (2003- )
- *Annales de l'institut Fourier* (Grenoble : 2001- )
- *Annales mathématiques Blaise Pascal* (Clermont-Ferrand : 2003- )
- *Journal de théorie des nombres de Bordeaux* (2004- )
- *MathS In Action* (SMAI : 2008- )
- *Actes des rencontres du CIRM* (2009- )
- Seminarios : JEDP (Evian : 2001- ), X-EDP (Paris : 1997- ), TSG (Grenoble : 2005-)

## *Fondos particulares*

**Archivo Nicolas Bourbaki** Documentos privados y manuscritos de este matemático famoso, escaneados por la difunta unidad CNRS « Archives de la création mathématique ».

En línea en <http://portail.mathdoc.fr/archives-bourbaki/>

**Archivo Laurent Schwartz** Documentos privados y manuscritos de este matemático famoso, procedente del anterior, escaneados por la biblioteca de la escuela Polytechnique (work-in-progress)

**Obras de Laurent Schwartz** escaneadas para la edición de un libro de obras escondidas con CD-ROM (SMF/X/CMD).

## Buscadores

- Gallica-Math** Portal de acceso mejorado a las colecciones matemáticas de Gallica, gracias a un catálogo de los artículos producido por MathDoc (Gallica-Math)
- Obras de numerosos matemáticos (oc)
  - *Journal de mathématiques pures et appliquées* (1836-1934) (Liouville)
  - *Répertoire bibliographique des sciences mathématiques* (1894-1912) (RBSM)
- LiNum** Catálogo (biblioteca virtual) de libros escaneados por otras instituciones (Gallica, Michigan U, Cornell, Göttingen, etc.) (LiNuM)
- mini-DML** Catálogo (biblioteca virtual) de artículos digitales producidos por otras instituciones (mini-DML)  
Prefiguración minimal de la DML

# *La biblioteca virtual de matemáticas*

## *Visión ( $\simeq$ Cornell, 2002)*

« A reference digital mathematics library should assemble as much as possible of the digital mathematical corpus in order to

- **preserve** it over the long term,
- make it **available online**
- at **reasonable cost**,
- in the form of an **authoritative** and **enduring** digital collection,
- **growing** continuously with publisher supplied new content,
- **augmented** with sophisticated search interfaces and interoperability services,
- developed and curated by a network of **institutions** »

# La biblioteca virtual de matemáticas

## Historia

- John Ewing. “Twenty Centuries of Mathematics: Digitizing and Disseminating the Past Mathematical Literature”. *Notices of the AMS*, 49(7):771–777, August 2002.
- Digital Mathematics Library. NSF planning project (2002–2003, Cornell University Library) “toward the establishment of a comprehensive, international, distributed collection of digital information and published knowledge in mathematics”.
- Mathematical Knowledge Management meetings (2001– ) + DML workshops (2008– ): technical challenges. Proyectos europeos EULER, LIMES, MowGLI. . .
- EMS’ EoI to the European Commission (2003), supported pilot implementation proposals to EC programmes (2003–2008: FP6, eContentplus. . . )
- AMS/MSRI proposal to the Moore foundation (2005)
- IMU support (2002–2006: Vision, Best practices)
- EuDML (=  $\frac{W}{2}$  DML) actualmente en negociación, prevista: 02/2010–01/2013

# *La biblioteca virtual de matemáticas*

## *Contenido corriente*

**America** JSTOR (260,000 textos), project Euclid (100,000), CMS (4,000)

**Asia** DML-JP (30,000 textos), China ??

**Europa** EuDML? (190,000 textos)

**Alemania** ERAM/JFM, GDZ, ELibM (85,000 textos)

**Francia** Gallica-Math, NUMDAM, CEDRAM, TEL (50,000 textos)

**Polonia** ICM/BWM (13,000 textos)

**Portugal** SPM/BNP (2,000 textos)

**España** DML-E (5,000 textos)

**Rep. Checa** DML-CZ (11,000 textos)

**Rusia** RusDML (13,000 textos)

**Bulgaria** BulDML (2,500 textos)

**Serbia** No formalised project (3,700 textos)

**Suiza** SwissDML (5,000 textos)

**Commercial** 700,000 textos?

**PYME** CUP 20 revistas, OUP 30, Hindawi 18, WdG 13, Wiley 42, T&F 58...

**Elsevier** 4 revistas en NUMDAM, 63 en Backfiles, 100 vivas (320,000 textos)

**Springer** 14 revistas en GDZ, 1+2 en NUMDAM, 120 en Online Archives,  
179 vivas (300,000 textos)

# *La biblioteca virtual de matemáticas*

## *Desafíos políticos*

- Renunciar a la visión inicial, centralizada y piramidal: crear una red horizontal de partners
- Crecer «bottom-up» más bien que «top-down»:  
local o nacional → regional o continental → mundial
- Encontrar modelos económicos bastante abiertos, y motivaciones para facilitar la colaboración de todas clases de partes afectadas

# *La biblioteca virtual de matemáticas*

## *Desafíos tecnológicos*

- Definir un esquema de metadatos bastante rico para funcionalidades útiles, pero razonable de producir para todos
- Desarrollar sistemas de producción automatizado de tales metadatos «bastante ricos» (GDZ, DML-CZ)
- Por cada texto, aprovechar todas las informaciones existentes (MR, ZM, y otras bases de datos)
- Los textos están escritos en múltiples idiomas, no hay idioma común en los metadatos
- Antiguos (escaneados) y nuevos (nacido digitales) textos están en ficheros gráficos (PDF), el texto puro puede servir de metadato, pero las formulas matemáticas mismas no son accesibles: deberíamos aprovechar de ellas





# The European Digital Mathematics Library

## CIP-ICT-PSP.2009.2.4 Open access to scientific information

- Consorcio** 13 participantes europeos (IST[P], CMD[F], UB[UK], FIZ/ZM[DE], MU[CZ], ICM[PL], CSIC[ES], EDPS[F], USC[ES], IMI-BAS[BG], IMAS[CZ], CSIU[GR], MML[UK])  
3 instituciones asociadas (EMS, SUBG, RusDML)
- Objetivos** Implementación piloto (para usuarios finales) de un punto de acceso único al contenido matemático de 11 partners; con muchas innovaciones
- Perfil** Duración: 3 años (02/2010-01/2013),  
487 PM, gasto global : 3,20 M€ (financiado por medio por la CE).
- Contenidos** 170,000 textos; 2 600 000 páginas  
**Retrodigitalizado** NUMDAM, Gallica, BWM, GDZ, SPM/BNP, DML-CZ, DML-E, RusDML.  
**Nacido digital** CEDRAM, ICM, EDPS, ELibM, DML-CZ, DML-E



# *The European Digital Mathematics Library*

## *Work plan*

- WP1 Project management
- WP2 Policies, exploitation, and dissemination
- WP3 Content aggregation
- WP4 System architecture and design
- WP5 Metadata repository and search engine implementation
- WP6 Web and service interface implementation
- WP7 Metadata enhancer implementation
- WP8 Association analyser implementation
- WP9 Annotation component implementation
- WP10 Accessibility component implementation
- WP11 Assessment and evaluation.

# *Gracias*

Thierry BOUCHE

Cellule MathDoc

BP 74

38402 Saint-Martin d'Hères

France

[thierry.bouche@ujf-grenoble.fr](mailto:thierry.bouche@ujf-grenoble.fr)

<http://www-fourier.ujf-grenoble.fr/~bouche/>

**Artículos sobre el tema:**

<http://www-fourier.ujf-grenoble.fr/~bouche/publist-dml.html>

**Presentaciones sobre el tema:**

<http://www-fourier.ujf-grenoble.fr/~bouche/radio-retro-in-actif.html#CT>